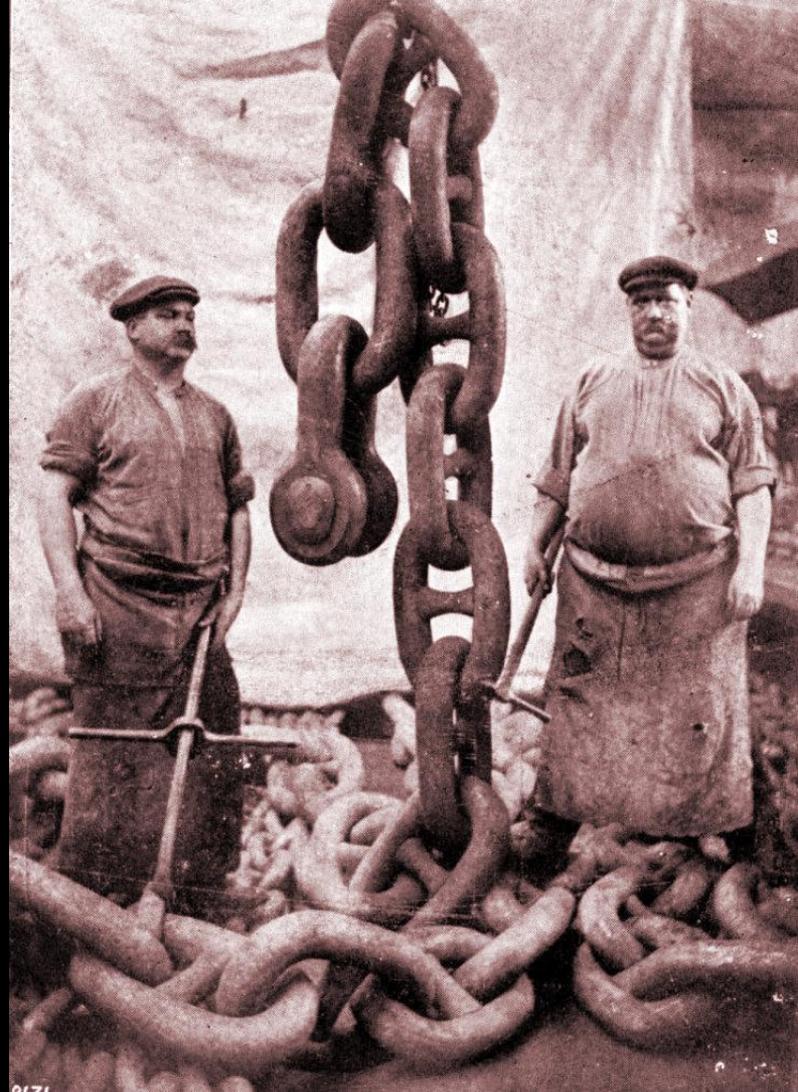


Building the sum of all human citations

Dario Taraborelli

@readermeter



How a collaboratively created, free online encyclopedia can accelerate access to expert-vetted knowledge.



“Wikipedia is not the bottom layer of authority, nor the top, but in fact the highest layer without formal vetting. In this unique role, it serves as **an ideal bridge between the validated and unvalidated Web.**”

Casper Grathwohl
[*Chronicle of Higher Education*](#)



The disappearance of provenance

Outline

1. The disappearance of provenance
2. A central repository of bibliographic and citation data
3. Introducing Wikiproject Source Metadata



Goat



"Billy goat" redirects here. For the Lance Corporal in the Royal Welsh, see *William Windsor (goat)*. For the band, see *Billy Goat (band)*.

This article is about the domesticated species. For other species, see *Capra (genus)*. For other uses, see *Goat (disambiguation)*.

The **domestic goat** (*Capra aegagrus hircus*) is a subspecies of goat domesticated from the wild goat of southwest Asia and Eastern Europe.

The goat is a member of the family *Bovidae* and is closely related to the sheep as both are in the goat-antelope subfamily *Caprinae*. There are over 300 distinct breeds of goat.^[1] Goats are one of the oldest domesticated species, and have been used for their milk, meat, hair, and skins over much of the world.^[2] In 2011, there were more than 924 million live goats around the globe, according to the *UN Food and Agriculture Organization*.^[3]

Female goats are referred to as "does" or "nannies", intact males as "bucks", "billies", or "rams" and their offspring are "kids". Castrated males are "wethers". Goat meat from younger animals is called "kid" or *cabrito* (Spanish), and from older animals is simply known as "goat" or sometimes called *chevon* (French), or in some areas "mutton" (which more often refers to adult sheep meat).

https://en.wikipedia.org/wiki/Goat#Life_expectancy

Domestic goat

Temporal range: .01–0 Ma

PreЄ Є O S D C P T J K P N

Neolithic – Recent g



A pygmy goat on a stump.

Conservation status

Goat



Life expectancy

Life expectancy for goats is between fifteen and eighteen years.^[29] An instance of a goat reaching the age of 24 has been reported.^[30]

Several factors can reduce this average expectancy; problems during kidding can lower a doe's expected life span to ten or eleven, and stresses of going into rut can lower a buck's expected life span to eight to ten years.^[30]

https://en.wikipedia.org/wiki/Goat#Life_expectancy



average goat lifespan

Web

Shopping

Images

Videos

News

More ▾

Search tools

About 367,000 results (0.39 seconds)

15 – 18 y

Goat, Lifespan



Feedback

Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion

Xin Luna Dong^{*}, Evgeniy Gabrilovich, Jeremy Heitz, Wilko Horn, Ni Lao,
Kevin Murphy[†], Thomas Strohmann, Shaohua Sun, Wei Zhang

Google, 1600 Amphitheatre Parkway, Mountain View, CA 94043
{lunadong|gabr|geremy|wilko|nlao|kpmurphy|tstrohmann|sunsh|weizh}@google.com

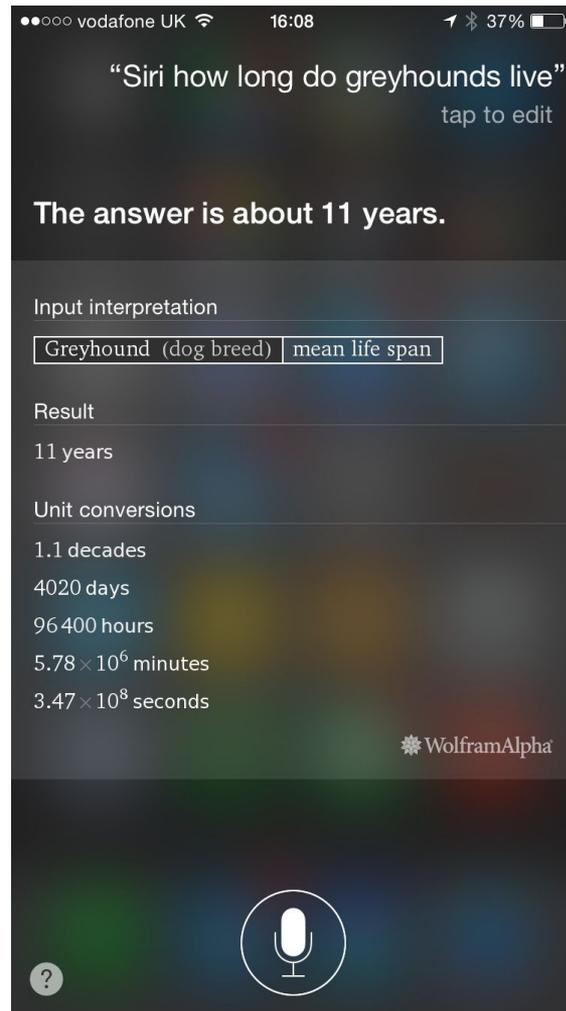
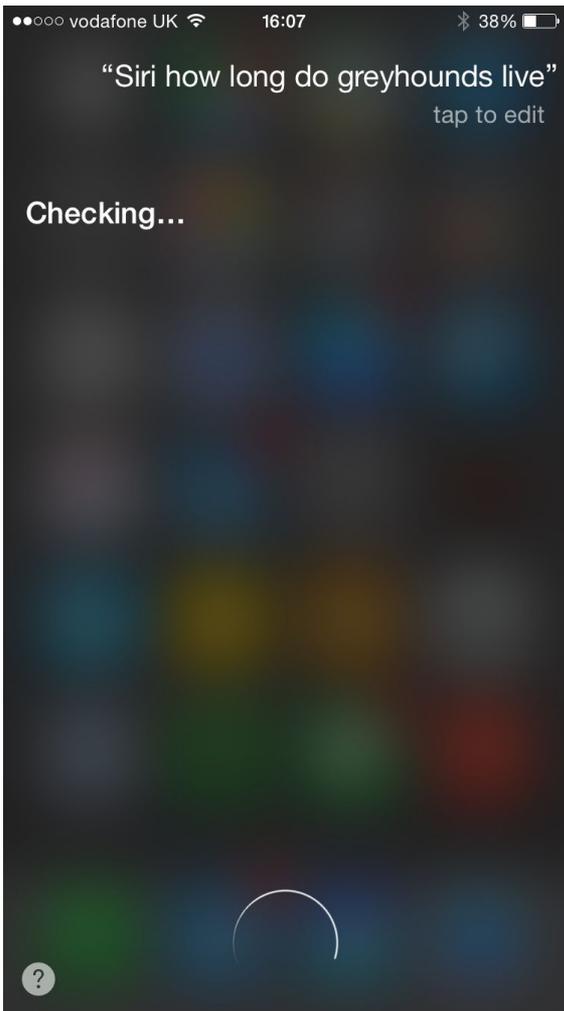
ABSTRACT

Recent years have witnessed a proliferation of large-scale knowledge bases, including Wikipedia, Freebase, YAGO, Microsoft's Satori, and Google's Knowledge Graph. To increase the scale even further, we need to explore automatic methods for constructing knowledge bases. Previous approaches have primarily focused on text-based extraction, which can be very noisy. Here we introduce Knowledge Vault, a Web-scale probabilistic knowledge base that combines extractions from Web content (obtained via analysis of text, tabular data, page structure, and human annotations) with prior knowledge derived from existing knowledge repositories. We employ supervised machine learning methods for fusing these distinct information sources. The Knowledge Vault is substantially bigger than any previously published structured knowledge repository, and features a probabilistic inference system that computes calibrated probabilities of fact correctness. We report the results of multiple studies that explore the relative utility of the different information sources and extraction methods.

In recent years, several large-scale knowledge bases (KBs) have been constructed, including academic projects such as YAGO [39], NELL [8], DBpedia [3], and Elementary/ DeepDive [32], as well as commercial projects, such as those by Microsoft¹, Google², Facebook³, Walmart [9], and others. (See Section 7 for a detailed discussion of related work.) These knowledge repositories store millions of facts about the world, such as information about people, places and things (generically referred to as entities).

Despite their seemingly large size, these repositories are still far from complete. For example, consider Freebase, the largest open-source knowledge base [4]. 71% of people in Freebase have no known place of birth, and 75% have no known nationality⁴. Furthermore, coverage for less common relations/predicates can be even lower.

Previous approaches for building knowledge bases primarily relied on direct contributions from human volunteers as well as integration of existing repositories of structured knowledge (e.g., Wikipedia infoboxes). However, these methods are more likely to yield head content, namely, frequently mentioned properties of frequently mentioned entities. Suh



External source:

▼ Dog breed data

- American Kennel Club. *American Kennel Club.* »
- Bryntel Technologies. *The Furry Critter Network.* »
- Collective Wisdom. *Best Dog Breeds.* »
- Czechoslovakian *Prinos.* »
- Dog Breed Info Co. *Dog Breed Info Co.* »
- Dog Breeds A-Z. *Dog Breeds A-Z.* »
- dogs-info.net. "http://www.dogs-info.net". »
- dogsindepth.com. *Dogs in Depth.* »
- Euro Puppy, Inc. *Euro Puppy.* »
- Erickson, L. *Viking Dog Breeds.* »
- FindTheBest.com. *FindTheBest.com.* »
- Great Breeds. *GreatBreeds.com.* »
- JCOM Designs, LLC. *TerrificPets.com.* »
- *longlivepuppies.com.* »
- Mitchell, P. *Canada's Guide to Dogs.* »
- My Lovely Pups. *mylovelypups.com.* »
- North American Kai Association. *kai-ken.org.* »
- *Pet Lovers World.* »

▼ Note

This list is intended as a guide to sources of further information. The inclusion of an item in this list does not necessarily mean that its content was used as the basis for any specific Wolfram|Alpha result.

Requests by researchers for detailed information on the sources for individual Wolfram|Alpha results can be directed [here](#).



Linking is a small act of generosity that sends people away from your site to some other that you think shows the world in a way worth considering. [...]

[Sources] that are not generous with linking [...] are **a stopping point in the ecology of information**. That's the operational definition of authority: The last place you visit when you're looking for an answer. If you are satisfied with the answer, you stop your pursuit of it. Take the links out and you think you look like more of an authority.

D. Weinberger (2012) Linking is a public good

<http://www.hyperorg.com/blogger/2012/02/26/2b2k-linking-is-a-public-good/>

BOSTON DAILY ADVERTISER

ALMANAC
FOR

1875

BEAUTIFULLY ILLUSTRATED.

PUBLISHED AT

29 COURT STREET, BOSTON,
MASS.



Wikipedia:Verifiability, not truth

From Wikipedia, the free encyclopedia



This [essay](#) contains the [advice](#) or opinions of one or more Wikipedia contributors on the [Verifiability](#) policies. Essays are *not* [Wikipedia policies or guidelines](#). Some essays represent widespread norms; others only represent minority viewpoints.

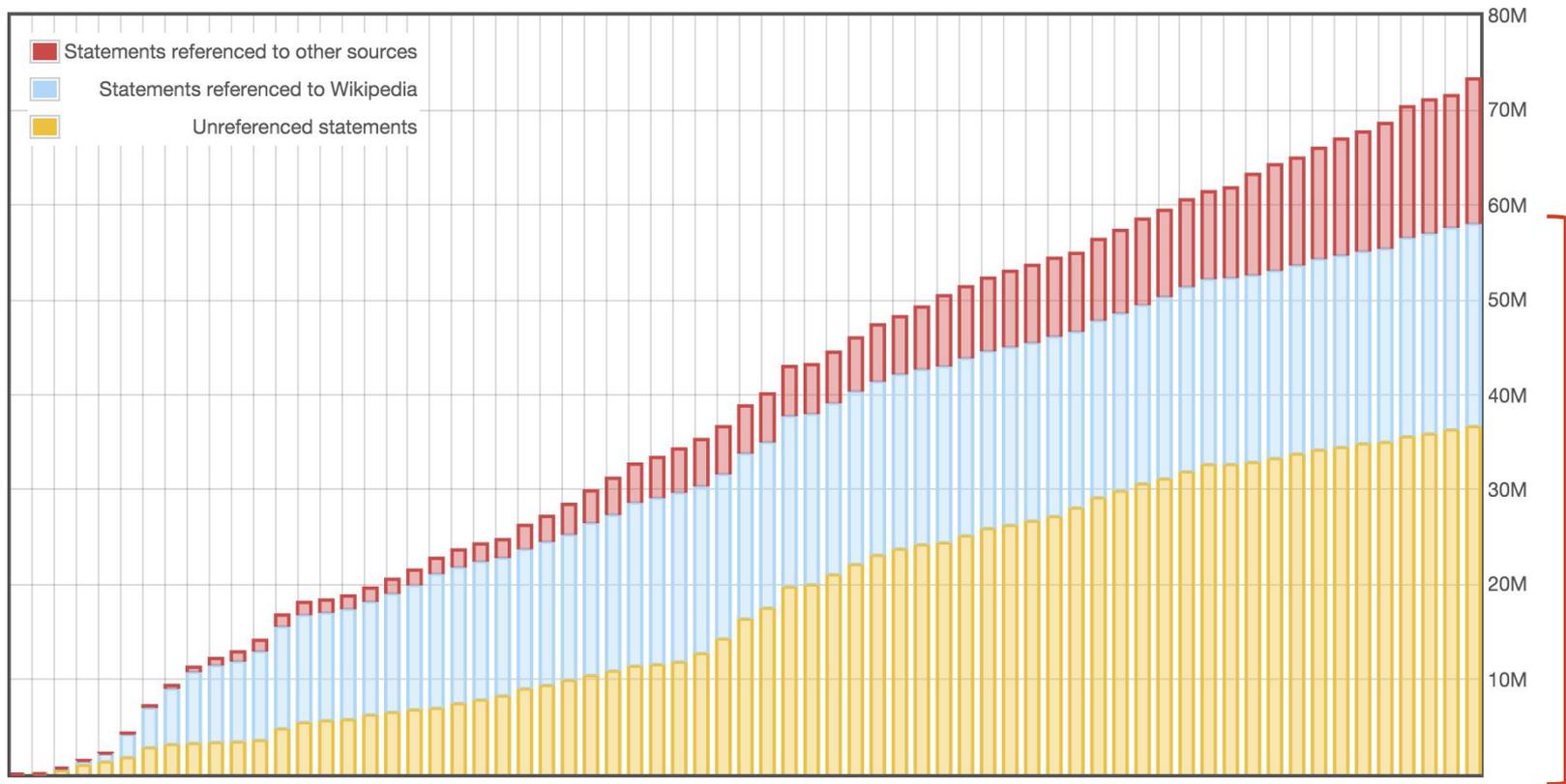
Shortcuts:
[WP:NOTTRUTH](#)
[WP:TRUTH](#)
[WP:VNT](#)



This page in a nutshell: Any material added to Wikipedia must have been published previously by a [reliable source](#). Editors may not add content solely because they believe it is true, nor delete content they believe to be untrue, unless they have verified beforehand with a [reliable source](#).

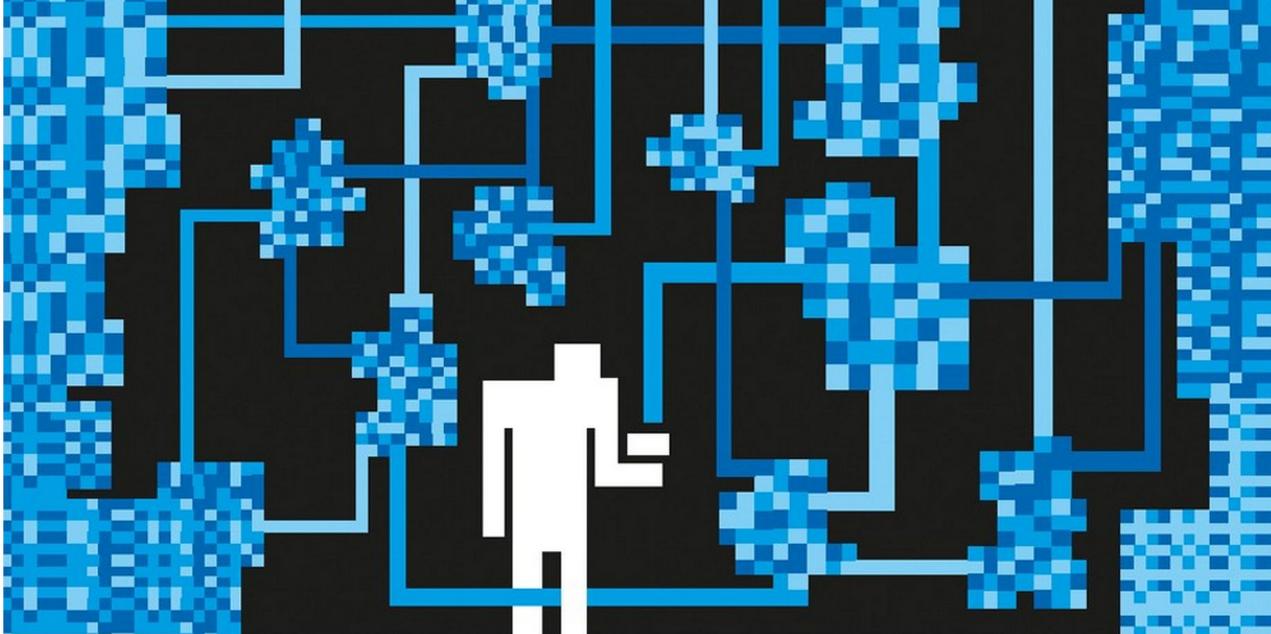
Wikipedia's core sourcing policy, [Wikipedia:Verifiability](#), used to define the threshold for inclusion in Wikipedia as "**verifiability, not truth**". "Verifiability" was used in this context to mean that material added to Wikipedia must have been published previously by a [reliable source](#). Editors may not add their own views to articles simply because they believe them to be correct, and may not remove sources' views from articles simply because they disagree with them.

The phrase "the threshold for inclusion is verifiability, not truth" meant that verifiability is a necessary condition (a minimum requirement) for the inclusion of material, though it is not a sufficient condition (it may not be enough). Sources must also be appropriate, and must be used carefully, and must be balanced relative to other sources per [Wikipedia's policy on due and undue weight](#).



80%

Building the sum of all human citations



Francisco Alandre

I Dreamed of a Perfect Database

How we order information can reshape our world.

BY **PAUL FORD** | December 2, 2015

<https://newrepublic.com/article/124425/dreamed-perfect-database>

Open Bibliography and Open Bibliographic Data

*Open Bibliographic Data Working
Group of the Open Knowledge
Foundation*



“For society to reap the full benefits from bibliographic endeavours, it is imperative that bibliographic data be made open — that is available for anyone to use and re-use freely for any purpose.”

<http://openbiblio.net/principles/>

Related projects [edit]

and historical discussion

Over the last six years, there have been a number of proposals for [new Wikimedia projects](#) and for technological changes to Mediawiki that would provide spaces for summaries, metadata, and discussion of references cited in Wikipedia articles and in other Wikimedia projects. These proposals include:

- [Wikicite \(metadata proposal\)](#) - a 2005 proposal to build an "expansion" to Mediawiki that would store and track bibliographic information for citations used in Wikipedia articles
- [Wikicat](#), [Wikicite \(2006 proposal\)](#), and [WikiProject Wikicite](#) - a series of related [new project proposals](#) based on the earlier [Wikicite metadata proposal](#)
- [Wikitextrose](#), a model for capturing citations as they are made on other Projects, and for recording the Type of citations [evidentiary, supporting, refuting, &c], intended for use with the above projects.
- [WikiScholar](#) - a [new project proposal](#) for a [Semantic Mediawiki](#) based bibliography



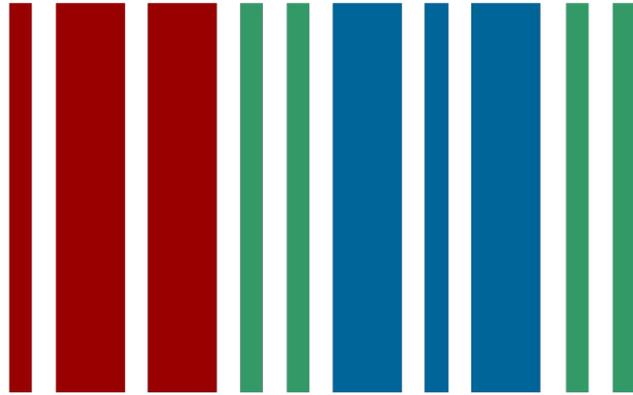
Outside of Wikimedia, [AcaWiki](#) is an active non-Wikimedia CC BY-licensed wiki that hosts summaries of scholarly articles along with metadata on scholarly works. AcaWiki current hosts about 800 summaries from over 100 contributors.

Other related Wikimedia projects whose goals would be served by a new project proving information citations include:

- [Research:Oral Citations](#) - a research project offering a rough proposal for building spaces to discuss oral citations in Wikipedia articles
- ['Fact and Reference Check'](#) ([en:wp](#))

<https://meta.wikimedia.org/wiki/Wikicite>

The sum of all human citations



WIKIDATA

Vision
Technology
Community
Scale
Licensing
Independence

The sum of all human citations

INTERNAL INTEGRATION



EXTERNAL INTEGRATION



ALTMETRICS SERVICES



SCHOLARLY DATA PROVIDERS



WikiProject Source MetaData

https://www.wikidata.org/wiki/Wikidata:WikiProject_Source_MetaData

1. Source metadata as items

Properties [\[edit\]](#)

Wikidata properties related to bibliographic metadata (Q20745594)	
Journal articles	author · editor · title · subtitle · publication date · published in · volume · issue · page · license · full text available at · reference URL · DOI · PubMed ID · PMCID · arXiv ID · ZooBank publication ID · language of work (or name) · original language of work · main subject · Commons category
Journals	title · publisher · ISSN · place of publication · editor · license · language of work (or name) · original language of work · part of · distribution · inception · dissolved or abolished · printed by · ZDB identifier · main subject · Commons category
Books (more)	author · translator · editor · illustrator · publisher · title · subtitle · license · full text available at · ISBN-10 · ISBN-13 · DOI · language of work (or name) · original language of work · main subject · Commons category
Authors	affiliation · e-mail · field of work · ORCID · VIAF identifier · ZooBank author ID · Commons category
Publishers	DOI Prefix · headquarters location · field of work · Commons category

See also: **WikiProject Source MetaData**, [WikiProject Periodicals](#), [WikiProject Books](#)

https://www.wikidata.org/wiki/Wikidata:WikiProject_Source_MetaData/Bibliographic_metadata_for_scholarly_articles_in_Wikidata

Ecological Guild Evolution and the Discovery of the World's Smallest Vertebrate (Q15567682)

scholarly article

Other properties	
published in	PLOS ONE peer-reviewed open-access journal for original scientific research
title	Ecological Guild Evolution and the Discovery of the World's Smallest Vertebrate [en]
license	CC BY 2.5 version of the free content license(s) from Creative Commons
instance of	scientific article article in an academic publication, usually peer reviewed
original language of work	English West Germanic language originating in England
language of work (or name)	English West Germanic language originating in England
author	Eric N. Rittmeyer researcher
	Allen Allison researcher
	Michael C. Gründler researcher
	Derrick K. Thompson researcher
	Christopher C. Austin researcher
publication date	2012-01-11
main subject	Paedophryne swiftorum species of amphibian
	Paedophryne amauensis a frog in the microhylid genus Paedophryne



External sources

DOI	10.1371/journal.pone.0029797
PMCID	3256195
PubMed	22253785
ZooBank publication	CC7DC93E-9BB6-4F1B-B96F-FD929E6FE1FD

Wikimedia projects

[Concept cloud](#)

<https://www.wikidata.org/wiki/Q15567682>

2. Cite by reference

The molecular origins of insulin go at least as far back as the simplest unicellular [[eukaryotes]].<ref name='LeRoith'>{{cite journal | vauthors = LeRoith D, Shiloach J, Heffron R, Rubinovitz C, Tanenbaum R, Roth J | title = Insulin-related material in microbes: similarities and differences from mammalian insulins | journal = Can. J. Biochem. Cell Biol. | volume = 63 | issue = 8 | pages = 839-49 | year = 1985 | pmid = 3933801 | doi = 10.1139/o85-106 }}</ref> Apart from animals, insulin-like proteins are also known to exist in Fungi and Protista kingdoms.

2. Cite by reference

The following code

```
{{#invoke:Cite
  | cite
  | item=Q15567682
}}
```

creates the reference string based on the item given with *qid*. For [Q15567682](#), this results in

[Eric N. Rittmeyer](#), [Allen Allison](#), [Michael C. Gründler](#), [Derrick K. Thompson](#) and [Christopher C. Austin](#), "Ecological Guild Evolution and the Discovery of the World's Smallest Vertebrate", *[PLOS ONE](#)*, 11 January 2012, doi: [10.1371/journal.pone.0029797](https://doi.org/10.1371/journal.pone.0029797) [↗](#), [CC BY 2.5](#)

3. Hold all the mappings

Scholarly article

QID ↔ DOI ↔ PMID ↔ PMC ↔ ARXIV ID

Author

QID ↔ ORCID ↔ VIAF ↔ Google Scholar ID

4. Map and measure source usage

Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid

Overview of attention for article published in Nature, April 1953



466

About this score

In the top 5% of all articles scored by Altmetric

SUMMARY News Blogs Twitter Facebook Wikipedia

? So far, Altmetric has seen 30 citations on 30 Wikipedia pages.

W ENGLISH **Gene**
Cited by user **Opabinia regalis** on **16 Apr 2015**
A gene is the molecular unit of heredity of a living organism. The word is used extensively by the scientific community for stretches of deoxyribonucleic acids (DNA) and...

W ENGLISH **Alex Stokes**
Cited by user **Chhandama** on **05 Aug 2014**
Alexander (Alec) Rawson Stokes (27 June 1919 – 5 February 2003) was a co-author of the second of the three papers published sequentially in Nature on 25 April 1953...

<https://www.altmetric.com/details/422759/wikipedia>

4. Link concepts and sources

Ecological guild evolution and the discovery of the world's smallest vertebrate [Q15567682]

main subject Paedophryne swiftorum [Q2169503]

↳ *parent taxon* Paedophryne [Q132975]

author Eric N. Rittmeyer [Q20745424]

↳ *affiliation* Louisiana Museum of Natural History [Q15958804]

↳ *location* Baton Rouge [Q28218]

5. Link and annotate sources

A *is licensed as* CC BY

A *is paywalled*

A *was retracted*

A *cites* B

A *extends* B

A *uses method in* B

A *uses data from* B

A *disagrees with* B

6. Research

“All pharmacology publications from the 1990s that have been retracted”

“All facts backed by works of physicists who graduated at Cambridge University in the 1980s”

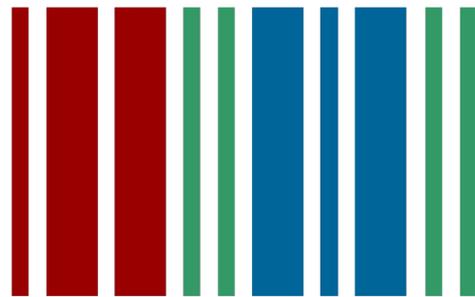
“All statements supported by articles published in the New York Times”

A provenance-preserving answer engine



The sum of all **human knowledge**

+



WIKIDATA

The sum of all **data and sources** backing human knowledge

What's next

- Run a pilot to populate source metadata for a specific discipline
- Design strategies for automatically importing source metadata and cross-linking the corresponding statements
- Refine the source data model, entity disambiguation
- Design a scalable human curation workflows to verify the quality of source metadata and the relevance of their use.

Questions?



dario@wikimedia.org

[@readermeter](https://twitter.com/readermeter)

[@wikiresearch](https://twitter.com/wikiresearch)

Image credits

Anchor Men - Mauretania

https://www.flickr.com/photos/twm_news/5161439712

CC0

King Billy of Dalkey Island

<https://www.flickr.com/photos/paulodonnell/5937678226>

CC BY

Boston Daily Advertiser Almanac

https://commons.wikimedia.org/wiki/File:1875_BostonDailyAdvertiser_Almanac.jpg

CC0

Angora Ram

<https://www.flickr.com/photos/roseannadana/3759787744>

CC BY ND

screenshots included under fair use